**ACCESS TO STATE GOVERNMENT INFORMATION INITIATIVE**
**EXTENDED ANNOTATED BIBLIOGRAPHY**
**June 8, 2004**
**Items labeled with a ** have been added since December 25**

**Access/Search**

**Jackson, Joab. "Agency Taxonomies Are a Tall Order, Experts Say." *Government Computer Network*, January 21 2004.
http://www.gcn.com/cgi-bin/udt/im.display.printable?client.id=gcndaily2&story.id=24669

Brief article about how all federal agencies need to develop a taxonomy of their websites, as mandated by the 2002 e-government act. Authors considers this a tough thing to do.

Bergman, Michael. "The Deep Web: Surfacing Hidden Values." In *Technology White Papers*: Bright Planet, 2001.
http://www.brightplanet.com/technology/deepweb.asp

Information on the Deep Web, the portion of the World Wide Web created by databases, and not indexed by conventional search engines (sometimes known as the invisible web, but the author does not like than term). Written with a slant to promote Bright Planet searching, but still provides useful information.

Feldman, Susan, and Chris Sherman. "The High Cost of Not Finding Information: An IDC White Paper." 10: IDC, 2001.
http://monkey.biz/Content/Default/Support/Resources/IDC_TheHighCostOfNotFindingInformation_1510.pdf

IDC places a dollar figure on the loss of businesses when their knowledge workers do not finding what they are looking for - believe that Fortune 100 companies waste $2.5 billion each year on fruitless searches. Paper specifically focuses on corporate intranets, but the conclusions can be extrapolated elsewhere.

*Invisible Web: What It Is, Why It Exists, How to Find It, and Its Inherent Ambiguity*. University of California, Berkeley, August 28, 2003.
http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html

Definitions of the "visible" and "invisible" web and how to find information from the invisible web. Some authors prefer the terms, "surface" and "deep" web.

Lyman, Peter, and Hal R. Varian. "How Much Information 2003?" Berkeley: University of California, Berkeley, 2003.
http://www.sims.berkeley.edu/research/projects/how-much-info-2003/

Most recent version of a study done at the University of California, Berkeley's School of Information Management Systems. The study analyses all information media - print, film, magnetic, and optical - and information flows - telephone, radio and TV, and the Internet - that came out in the year 2002 and attempts to quantify the results. These results are compared to the results of the 2000 study, and estimate a 30% increase in information. Take a look at the executive summary, which summarizes all the information in tables. It is hard to say how much of the information is unique versus how much duplicates other information.

**Woods, William A. "Searching Vs. Finding." *ACM Queue* 2, no. 2 (2004).
http://www.acmqueue.com/modules.php?name=Content&pa=showpage&pid=137

Description of the difficulties of searching and actually finding useful information. This article looks at two major reasons search results are less than ideal--morphological relationships and semantic relationships--and explores search options that try to take these issues into account, including stemming, taxonomies, and specific passage retrieval. Not all search mechanisms will work as well in all situations, and the author specifically believes that expanding search results through stemming and passage retrieval may do more harm than good for a general web search, given the size of the index and the diverse nature of the information. A good introduction to the technical details of information retrieval.

## Digital Libraries

**Conway, Paul. "Deep Infrastructure Supports Digital Library Services." *Syllabus Magazine*, May 2004.
http://www.syllabus.com/article.asp?id=9362

Description of the planned digital library at Duke University, and the infrastructure needed to support the library. The digital library will be successful by combining the efforts occurring across campus to develop a deep infrastructure using the expertise in the Duke Library Information Technology Services and campus IT. Three systems will be part of the digital library: Ex Libris, to provide a metadata repository and suite of search and discovery tools; Zope Corporation's web content management system; and DSpace for institutional archiving.

Cruse, Patricia, Chuck Eckman, John Kunze, and Heather Christenson. "Web-Based Government Information: Evaluating Solutions for Capture, Curation, and Preservation." California Digital Library, 2003.
http://www.cdlib.org/programs/Web-based_archiving_mellon_Final.pdf

Report of the work done by the California Digital Library to capture and preservation digital government documents. Includes findings from a survey of digital web archive initiatives.

**Government Information**

Adler, Prudence S. "Rethinking the Federal Depository Library Program." *ARL Bimonthly Report*, August 2003.
http://www.arl.org/newsltr/229/fdlp.html

Summarizes the Depository Library Council meeting in April, 2003, and how the GPO is rethinking its services in the wake of the digital age. States that "95 % of government information will be digital in 5 years." Discusses different delivery methods and public access.

British Library. "Historic Change in Legal Deposit Law Saves Electronic Publications for Future Generations - Bill to Extend Legal Deposit to UK Non-Print Materials Receives Royal Assent." 2003.
http://www.bl.uk/cgi-bin/press.cgi?story=1382

Press Release describing the new legislation that extends previous legal deposit legislation passed nearly 100 years ago in 1911. The Act enshrines the principle that electronic or e-publications and other non-print materials will be deposited in the future under secondary legislation. In ensures that these publications can be saved as part of the published archive and become an important resource for future generations of researchers and scholars. Description of the materials that will now be included in depository law.

**Categorization of Government Information Working Group. "Defining What Government Information Is to Be Categorized (Draft)." U.S. Federal Interagency Committee on Government Information, May 13 2004.
http://www.gpoaccess.gov/cgiwg/pdf/cgiwgroup/revMay2004.pdf

Draft report that is part of the work of the Categorization of Government Information Working Group. The document outlines a definition of what types of U.S. Federal Government information are in scope for a standardized categorization schema. The draft definition agreed upon for now is "'categorizable Government information' means any information product, regardless of form or format, that an agency discloses, publishes, disseminates, or makes available to the public, as well as information produced for administrative or operational purposes that is of public interest or education value." The first draft is still available at http://www.gpoaccess.gov/cgiwg/pdf/cgiwgroup.pdf.

**Chase, Eleanor, Beth Clausen, Teffeni J. Fontno, Greta Marlatt, Amiee Piscitelli Quinn, and Earl Shumaker. "Federal Documents Task Force on Permanent Public Access Final Report." December 29, 2003.
http://tigger.uic.edu/~aquinn/access/publicaccessindex.html

The Task Force examined the policies of the current administration in the federal government, noting the decline in the availability of government information. They update and restate the GODORT (Government Documents Round Table) principles of the need for free and easy access to government information. The paper discusses the potential but failed attempt to transfer printing power away from the GPO, the disappearance of information from government websites, the censoring of paper documents, Bush's executive order freezing release of presidential papers, and attempts to privatize government services. The paper includes an extensive bibliography.

Drake, Miriam. "Agreement Ensures Permanent Public Online Access to Government Information." *Information Today NewsBreaks*, August 25 2003.
http://www.infotoday.com/newsbreaks/nb030825-1.shtml

Summary of the agreement between GPO and NARA that announced that NARA will take over responsibility for long-term preservation of federal digital information hosted on the GPO Access site.

Matthews, Richard J, Anne E. Burnett, Charlene C. Cain, Susan L. Dow, David L. McFadden, and Mary Alice Baish. *State-by-State Report on Permanent Public Access to Electronic Government Information*. Chicago, IL: American Association of Law Libraries, 2003.
http://www.ll.georgetown.edu/aallwash/State_PPAreport.htm

The AALL surveyed 50 states, D.C. and Puerto Rico to determine current preservation efforts for digital government information. Each state has a separate chapter describing the current situation and there is an overview of the whole survey at the beginning of the publication. Certain states' approaches, including North Carolina's, are highlighted, if they are considered particularly important or innovative.

**Milbank, Dana. "White House Web Scrubbing: Offending Comments on Iraq Disappear from Site." *Washington Post*, December 18 2003, A05.
http://www.washingtonpost.com/wp-dyn/articles/A9821-2003Dec17.html

Brief article describing changes the Bush Administration has made to government websites, done mainly to support his own political agenda. This has included removing information that contradicts later statements.

Sepic, Ron, and Kate Kase. "The National Biological Information Infrastructure as an E-Government Tool." *Government Information Quarterly* 19 (2002): 407-24.

Description of the National Biological Information Infrastructure - a web-based system that provides access to data and information on the nation's biological resources. An e-government initiative as well as an example of bringing to together digital government information from disparate sources. While the article is not available online, the website may be viewed at http://www.nbii.gov/.

**US Government Printing Office. "Collection of Last Resort (Draft)." April 6 2004.
http://www.access.gpo.gov/su_docs/fdlp/pubs/clr.pdf

Draft plan of the proposed "Collection of Last Resort" to be implemented by the Government Printing Office. GPO provides information about scope, collection, preservation, and access to the collection. Both digital and tangible objects will be collected, and tangible objects will be digitized for access. Discussion of the draft took place at the FDLP Council meeting in the Spring, with a final version due out later this year.

West, Darrell M. "Achieving E-Government for All: Highlights from a National Survey." Benton Foundation, New York State Forum of the Rockefeller Institute of Government, 2003.
http://www.benton.org/publibrary/egov/access2003.html

Reports the finding of a national survey concerned with the accessibility of government information and programs available through the Internet.

"When Government Invests in It, Benefits to Constituents Often Exceed Internal Benefits to Government, Says Study." *Government Technology*, November 2003.
http://www.govtech.net/news/news.phtml?docid=2003.10.10-72208

Brief release stating that traditional measures of return on investment do not measure the value of government IT projects, which can save constituents time and money by providing ready access to information and e-government.

**Libraries - General**

Keller, Michael, Victoria A. Reich, and Andrew C. Herkovic. "What Is a Library Anymore Anyway?" *First Monday* 8, no. 5 (2003).
http://www.firstmonday.org/issues/issue8_5/keller/index.html

Libraries in the future will undertake local control, especially for long-term preservation and accessibility of digital as well as analog collections. Failure to embrace that role would cause libraries and librarians rapidly to lose relevance and value as internet and other digital resources develop. Local control of collections is critical both to assure permanence and to provide a key degree of selectivity, which - contrary to the irrational exuberance of making everything available to everybody - is vital to providing service to communities of readers. Librarians need new tools, such as the LOCKSS system, to enable both persistence and selection of electronic information.

**Metadata**

**"A Dozen Primers on Standards." *Computers in Libraries* 24, no. 2 (2004): 18-28.
http://www.infotoday.com/cilmag/feb04/primers.shtml

One-page primers on standards designed to provide a quick over of the most important standards for exchanging information. Standards covered: ARK, DOI, METS, MODS, NCIP, OAI-PMH, ONIX, OpenURL, RDF, RSS, Shibboleth, and SRW-SRU. The standards are mainly for the exchange and organization of metadata, used by the library and publishing world. Information about each standard is written by an expert in the field and includes numerous links for more information. A place to go when acronyms get overwhelming.

**Caplan, Priscilla. *Metadata Fundamentals for All Librarians*. Chicago: American Library Association, 2003.

This book provides an excellent overview of what metadata is and how it is being used to organize digital information. The approach is definitely focused for librarians, coming from a long history of description and cataloging. The first part of the book describes the principles and practices of metadata in general, including the basics of what metadata is, how is it created and stored, controlled vocabularies and classification, interoperability, and uses on the web. The second part of the book focuses on common metadata schemes, including Dublin Core, GILS, and others. There are also chapters dealing with types of metadata in general, including administrative metadata, structural metadata, and rights metadata. Each chapter includes a list for further readings. Highly readable and highly informative. It is not available on the Internet.

Gilliland-Swetland, Anne J. "Setting the Stage." In *Introduction to Metadata: Pathways to Digital Information*, edited by Murtha Baca. Los Angeles: Getty Research Institute, 2000.
http://www.getty.edu/research/institute/standards/intrometadata/

Overview of what metadata is, the different types of metadata - Administrative, Descriptive, Preservation, Technical, Use - and why metadata is important.

**"Metadata Resources Guide." Edmonton, Alberta: Information Management Branch, Alberta Government Services, May, 2004.
http://www.im.gov.ab.ca/publications/pdf/MetadataResGuide.pdf

Bibliography of resources and links relating to all types of metadata - set up by county, schema, encoding format, and domain specific metadata. The guide is updated regularly.

*Mets: Metadata Encoding and Transmission Standard Official Web Site*. Library of Congress.
http://www.loc.gov/standards/mets/

Overview of METS, which is a metadata standard designed to bring together descriptive and administrative metadata, and provide ways to link multiple resources dealing with the same subject. It could be an oral history, available in streaming audio and transcription, or a webpage with many links. There are examples of METS in XML. The goal of METS is to allow for preservation and sharing of information between digital libraries.

**Metadata - GILS**

Stephenson, Francine, Joel Sigmon, and John Sublette. "Toward a Government Information Locator Service for North Carolina: A Project of the Office of State Planning." Raleigh, NC: Office of State Planning, 1997.
http://ncgils.state.nc.us/plan.html

Overview of NC GILS when it was first initiated.

**Metadata - Preservation**

**Lavoie, Brian F. "Implementing Metadata in Digital Preservation Systems: The PREMIS Activity." *D-Lib Magazine* 10, no. 4 (2004).
http://www.dlib.org/dlib/april04/lavoie/04lavoie.html

A brief summary of the work of the PREMIS working group, jointly sponsored by RLG and OCLC, focused on the topic of implementing preservation metadata within digital archiving systems. PREMIS is approximately halfway done with their two primary objectives: creating a core set of preservation data elements and a data dictionary to allow practical implementation of those elements. Full reports of the core elements and implementation strategies will be available once they are completed.

"Preservation Metadata for Digital Objects: A Review of the State of the Art: A White Paper by the OCLC/RLG Working Group on Preservation Metadata." 49. Dublin, OH: OCLC, 2001.
http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf

Provides a definition of preservation metadata; high level requirements for a broadly applicable, comprehensive preservation metadata framework; the Open Archival Information System (OAIS) reference model, a potential starting point for developing the preservation metadata framework; review and synthesis of existing preservation metadata approaches; and identification of points of convergence/divergence among existing approaches. Heavy, dense, and technical.

**Preservation**

*Building a National Strategy for Preservation: Issues in Digital Media Archiving*. Edited by Amy Friedlander. Washington, D.C.: Council on Library and Information Resources and the Library of Congress, 2002.
http://www.clir.org/pubs/abstract/pub106abst.html

Overview of different media and preservation challenges. This is included in the appendices for *Preserving Our Digital Heritage*. Of particular interest in *Archiving the World Wide Web* by Peter Lyman.

Conway, Paul. "Preservation in the Digital World." Council on Library and Information Resources, 1996.
http://www.clir.org/pubs/abstract/pub62.html

Seminal work detailing the difficulties of preserving digital information.

Hedstrom, Margaret, Seamus Ross, Kevin Ashley, Birte Christensen-Dalsgaard, Henry Gladney, Claude Huc, Anne R. Kenney, Reagan Moore, and Erich Neuhold. "Invest to Save: Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation." National Science Foundation, Network of Excellence for Digital Libraries, 2003.
http://delos-noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/digitalarchiving/Digitalarchiving.pdf

This report identifies research challenges, development opportunities and solutions for problems associated with the preservation of digital information. Consisting of six main areas, the report presents digital preservation challenges; looks at the benefits of long-term preservation; summarizes a set of principles and assumptions to guide the research agenda; highlights the work to date within this field; discusses a research agenda, including emerging research; and addresses the need for additional areas of support.

RLG-OCLC Working Group on Digital Archive Attributes. "Trusted Digital Repositories: Attributes and Responsibilities." In *RLG-OCLC Report*. Mountain View, CA: RLG, 2002.
http://www.rlg.org/longterm/repositories.pdf

Description of what makes up a trusted digital repository. Most useful are a checklist of attributes at the end and the appendix describing the Open Archival Information System

Smith, Abby. "Digital Preservation: An Individual Responsibility for Communal Scholarship." *EDUCAUSE Review*, May/June 2003, 10-11.
http://www.educause.edu/ir/library/pdf/erm0338.pdf

A description of the challenges associated with preserving digital information, set in terms for scholars to understand

Spedding, Vanessa. "Data Preservation: Great Data, but Will It Last?" *Research Information*, Spring 2003.
http://www.researchinformation.info/rispring03data.html

Brief overview of different preservation projects around the world, including the Open Archive Initiative (OAI), NDIIPP, DSpace, CAMiLEON.

Task Force on Archiving of Digital Information. "Preserving Digital Information." Commission on Preservation and Access and Research Libraries Group, 1996.
http://www.rlg.org/ArchTF/

Seminal work that really first brought to the attention to the entire library world the issues involved in digital preservation. Though the article is now 7 years old, many of its key points and recommended course of action still hold true today.

Warner, Dorothy. ""Why Do We Need to Keep This in Print? It's on the Web...": A Review of Electronic Archiving Issues and Problems." *Progressive Librarian*, no. 19-20 (2002).
http://libr.org/PL/19-20_Warner.html

Cautionary article about the dangers of relying on the Web for access to information, strong belief in libraries as custodians of information as well as access points. Discussion of the problems of digital preservation and overview of some of the initiatives nationally and globally, though GPO has since entered an agreement with NARA to preserve digital government publications.

Webb, Colin. "Barriers or Stepping Stones? Impediments to Digital Archiving and Preservation Programs." *CLIR Issues*, July/August 2003.
http://www.clir.org/pubs/issues/issues34.html

Author considers the stumbling blocks that the National Library of Australia faced when trying to start a preservation program: inability to determine where to start, lack of sufficient expertise, an absence of easily obtainable and trusted tools, and unrealistic expectations about costs.

**Preservation - Economics**

Chapman, Stephen. "Counting the Costs of Digital Preservation: Is Repository Storage Affordable?" *Journal of Digital Information* 4, no. 2 (2003).
http://jodi.ecs.soton.ac.uk/Articles/v04/i02/Chapman/

The Harvard University Library and the Online Computer Library Center, Inc. (OCLC) each manage centralized repositories optimized for long-term storage of library collections. Both organizations fully recover operational expenses by charging owners annual rates for managed storage services, regardless of materials use. The Harvard Depository assesses rates for analog storage per billable square foot. The OCLC Digital Archive assesses rates per gigabyte for storage of digital objects. Formats are significant, but not sole factors in determining preservation costs in these models. Owners' definitions of content integrity and tolerance for risk, which can change over time, are also important variables in the complex equation of preservation costs and affordability.

Lavoie, Brian F. "The Incentives to Preserve Digital Matierals: Roles Scenarios, and Economic Decision-Making." Dublin, OH: OCLC Online Computer Library Center, INC, 2003.
http://www.oclc.org/research/projects/digipres/incentives-dp.pdf

Economic issues are a principal component of the research agenda for digital preservation. Economics is fundamentally about incentives, so a study of the economics of digital preservation should begin with an examination of the incentives to preserve. Securing the long-term viability and accessability of digital materilas requires an appropriate allocation of incentives among key decision-makers in the digital preservation process. But the circumstances under which digital preservation takes place often lead to a misalignment of preservation objectives and incentives. Identifying circumstances where insufficient incentives to preserve are likely to prevail, and how this can be remedied, are necessary first steps in developing economically sustainable digital preservation activities.

**Preservation - Formats**

Darlington, Jeffrey. "Pronom--a Pratical Online Compendium of File Formats." *RLG DigiNews*, October 15 2003.
http://www.rlg.org/preserv/diginews/v7_n5_feature2.html

Description of a new service offered by the National Archives of the U.K., called PRONOM, which is a database of file formats and operating systems that support them. It is designed to help in emulation projects and help find migration paths, both of which can combat loss of information through software obsolesence.

Deare, Steven. "Australia's History Archives in Openoffice.Org." *PC World*, October 14 2003.
http://www.pcworld.idg.com.au/index.php?id=1991153367&fp=2&fpid=1

Description of how the National Archives of Australia plans to use OpenOffice.org as a method to view records in a non-proprietary format.

Emery, Gail Repsher. "E-Documents Need E-Preservation." *Washington Technology* 17, no. 23 (2003).
http://www.washingtontechnology.com/news/17_23/federal/20235-1.html

Discussion of the creation of PDF/A and plan to pursue ISO certification for it. Discussion of the implications for government information

**Jackson, Joab. "Feds Help Create PDF Archiving Standard." *Government Computer News*, May 19 2004.
http://www.gcn.com/vol1_no1/daily-updates/25986-1.html

Update on the status of PDF/A, which hopes to have a final standard out by the end of 2005. Brief article also describes some of the distinguishing features of PDF/A from PDF, including: embedded fonts, standardized metadata, multiple language support, and no scripts for executable programs within the document. It is designed for static documents.

LeFurgy, William G. "PDF-A: A New Preservation Format." *Government Record News*, May-June 2003.
http://www.archivists.org/saagroups/gov/newsletters/May2003.asp#digital

2 page summary of PDF-A, its advantages over traditional PDF, and the work toward making it an ISO standard. PDF-A will be a subset of PDF, completely self-contained, and use XMP for plain-text metadata markup.

**LeFurgy, William G. "PDF/A: Developing a File Format for Long-Term Preservation." *RLG DigiNews* 7, no. 6 (2003).
http://www.rlg.org/preserv/diginews/v7_n6_feature1.html

An short description of the proposed PDF/A standard as an archival format, specifying the distinctions between traditional PDF and PDF/A and why having the PDF/A standard is important for document presentation. Also provides an overview of why proprietary software is inappropriate for archival purposes.

Looney, Michael. "The Need for Digital Archiving Standards." *Syllabus: Technology for Higher Education*, March 2003.
http://www.syllabus.com/article.asp?id=7362

Discusses the importance of having standards for preserving digital information, and goes into detail about PDF/A

**Preservation - National Initiatives**

Beagrie, Neil. "National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity." Council of Library and Information Resources and Library of Congress, 2003. http://www.clir.org/pubs/abstract/pub116abst.html

The author surveyed national preservation programs at the countries listed in the title to help information the US on its own project, the NDIIPP. There are observations of principles trends and lessons for all countries, as well as specifics on each of the programs, their lessons learned, and the potential for collaboration.

**Hakala, Juha. "Archiving the Web: European Experiences." *Tietolinja*, no. 2 (2003). http://www.lib.helsinki.fi/tietolinja/0203/index.html

Description of the NEDLIB project, which developed a web harvester, and the Nordic Web Archive, which developed a search/access mechanism to the archive. The projects at a point of transition. The Finnish Center for Scientific Computer, which developed the NEBLIB harvester, no longer supports it, and the Nordic National Libraries are unsure whether they can continue to support the NWA tool set once funding runs out. A new project in 2003 has been started, with European national libraries and the Internet Archive, the International Internet Preservation Consortium to develop the next generation of tools to support Internet archiving.

Heslop, Helen, Simon Davis, and Andrew Wilson. "An Approach to the Preservation of Digital Records." Canberra: National Archives of Australia, 2002. http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper.pdf

Brief discussion of the Australian approach to preserving digital records. The report includes an overview of the problems and approach to digital preservation, description of the principles guiding the Australia National Archives, and a description of XML and how it will play into the preservation

*Keeping Government Publications Online: A Guide for Commonwealth Agencies*. National Library of Australia, National Archives of Australia, National Office for the Information Economy, 2002. http://www.nla.gov.au/guidelines/govpubs.pdf

An excellent brochure put forth by Australia's national information agencies about their program to provide permanent public access to Australian government documents. It details each institution's role and responsibilities, defines publication, and describes the steps agencies need to take to be in compliance with the program. It also has a brief reason for why this needs to be done as well.

**Library of Congress. "Update to the NDIIPP Architecture." Washington, D.C.: Library of Congress, 2004.
http://www.digitalpreservation.gov/repor/NDIIPP_v02.pdf

Draft of version 0.2 of the technical architecture for the National Digital Information Infrastructure and Preservation Program. There are several changes from the earlier version of the architecture (which was published in *Preserving our Digital Heritage)*. The four layers have been condensed to three, with the gateway layer removed. The remaining three layers have been renamed Upper (formerly Interface), Middle (formerly Collections), and Lower (formerly Repository). Earlier the infrastructure had specified that metadata should be stored completely separately from the object; now the infrastructure states that metadata can be stored with the object at the lower layer and also in the middle layer. The functions of the remaining three layers have also been more fully described. This a draft requesting comment from interested parties.

*Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program*. Washington, D.C.: Library of Congress, 2002.
http://www.digitalpreservation.gov/index.php?nav=3&subnav=1

Comprehensive plan for digital preservation put forth by the Library of Congress. This is the first step of the National Digital Information Infrastructure and Preservation Program

**United Kingdom Public Record Office. *UK Central Government Web Archive*. Public Record Office, 2004.
http://www.pro.gov.uk/webarchive/

The website for the new UK Central Government Web Archive, which became available to the public in February 2004. The archive was developed by the National Archives, using the services of the Internet Archive. A sample of 51 representative government websites are part of the archive. New snapshots are created every week or every 6 months, depending on the website. Visitors can search for older websites by entering the URL or by browsing by department.

"Where Websites Go to Die." *Sydney Morning Herald*, October 17 2003.
http://www.smh.com.au/articles/2003/10/16/1065917549444.html

Description of Australia's attempts to preserve older websites through the PANDORA program.

**Preservation - OAIS**

Lavoie, Brian F. "The Open Archival Information System Reference Model: Introductory Guide." In *DPC Technology Watch*. Dublin, Ohio: OCLC, 2004.
http://www.dpconline.org/graphics/reports/index.html#intoais

A brief introduction to the Open Archival Information System Reference Model, designed for people who are not otherwise familiar with the system. The article discusses the history of how the model came about, the system as a whole, the components of the functional model, the components of the information model, and other high-level work being done toward the preservation of digital information.

**Preservation - Repositories**

*DSpace Federation*. MIT Libraries
Hewlett-Packard Company, 2003.
http://www.dspace.org

Official website for DSpace, an institutional repository system, created by MIT, with Hewlett-Packard. Includes an overview of the system, articles about it, and links to MIT's website to see DSpace in action.

**Open Society Institute. *A Guide to Institutional Repository Software*. 2nd ed. New York, 2004.
http://www.soros.org/openaccess/pdf/OSI_Guide_to_Institutional_Repository_Software_v2.pdf

Brief description and table of comparisons of five open-source, OAI compliant, publicly available institutional repository applications: ARNO, CDSware, DSpace, Eprints, Fedora, i-Tor, and MyCoRe. All but DSpace and Fedora are European applications. Eprints currently enjoys widespread use as a journal repository. This updates an earlier edition produced by the Open Society Institute last year.

Smith, McKenzie, Mick Bass, Greg McClellan, Robert Tansley, Mary Barton, Margret Branschofsky, Dave Stuve, and Julie Harford Walker. "DSpace: An Open Source Dynamic Digital Repository." *D-Lib Magazine* 9, no. 1 (2003).
http://www.dlib.org/dlib/january03/smith/01smith.html

For the past two years the Massachusetts Institute of Technology (MIT) Libraries and Hewlett-Packard Labs have been collaborating on the development of an open source system called DSpace that functions as a repository for the digital research and educational material produced by members of a research university or organization. Running such an institutionally-based, multidisciplinary repository is increasingly seen as a natural role for the libraries and archives of research and teaching organizations. As their constituents produce increasing amounts of original material in digital form--much of which is never published by traditional mean--the repository becomes vital to protect the significant assets of the institution and its faculty. The first part of this article describes the DSpace system including its functionality and design, and its approach to various problems in digital library and archives design. The second part discusses the implementation of DSpace at MIT, plans for federating the system, and issues of sustainability.

Staples, Thornton, Ross Wayland, and Sandra Payette. "The Fedora Project: An Open-Source Digital Object Repository Management System." *D-Lib Magazine* 9, no. 4 (2003).
http://www.dlib.org/dlib/april03/staples/04staples.html

Description of Fedora, an open-source institutional digital repository developed by Cornell University.

**Wheatley, Paul. "Institutional Repositories in the Context of Digital Preservation." In *DPC Technology Watch Report 04-02*, 2004.
http://www.dpconline.org/docs/DPCTWf4word.pdf

Brief description of the different aspects of preservation functions that institutional repositories will have to address, including persistent identification, ingest, representation systems (mechanisms for storing and utilizing knowledge on how to gain access to the intellectual content of the digital object - e.g. file format registries), technology watch, rendering, overall repository structure, and recording change metadata. The report specifically addresses 5 open-source repository systems: CDSware, DSpace, EPrints, FEDORA, and MyCoRe. Only DSpace lists preservation as one of its fundamental goals, though developing systems not yet in use (and thus not reviewed by this report) are examining preservation concerns. The author believes that representation systems need the most research and development at this time.

**Preservation - Rescue**

Darlington, Jeffrey, Andy Finney, and Adrian Pearce. "Domesday Redux: The Rescue of the Bbc Domesday Project Videodiscs." *Ariadne*, no. 36 (2003).
http://www.ariadne.ac.uk/issue36/tna/

The BBC in 1986 created a highly sophisticated multimedia product called the Doomsday Project to provide a database of what Britain was like in 1986 (named after William the Conqueror's similar project in 1086 to do the same thing). The project was successful, but relied upon unusual formats which soon become obsolete. The CAMiLEON project took on the case in 2002 to restore and emulate the Doomsday project on modern-day equipment. The article describes the success of the preservation initiative and also some of the stumbling blocks. How authentic should the experience be, or can the data and images be cleaned up and improved to today's standards?

**Preservation - State Initiatives**

**"Washington State Digital Archives Project Feasibility Study." Washington Secretary of State, 2003.
http://www.secstate.wa.gov/archives/pdf/digital_archives/Feasibility%20Study.pdf

Description of the new digital archives for Washington State, to open in July 2004, in Cheney, Washington, the first state level archives of its kind. The study goes through the business needs that require a digital archives, a look at the alternatives to having a centralized digital archive, a description of the archives and how it will be implemented, and estimated costs and a proposed timeline.